Journal of Experimental Botany, Vol. 64, No. 7, pp. 2027–2040, 2013
doi:10.1093/jxb/ert070 Advance Access publication 27 March, 2013
This paper is available online free of all access charges (see http://jxb.oxfordjournals.org/open_access.html for further details)

Journal of Experimental Botany
www.jxb.oxfordjournals.org

RESEARCH PAPER

Novel insights into the composition, variation, organization, and expression of the low-molecular-weight glutenin subunit gene family in common wheat

Xiaofei Zhang[1,*], Dongcheng Liu[1,*], Jianghua Zhang[1,2], Wei Jiang[1], Guangbin Luo[1], Wenlong Yang[1], Jiazhu Sun[1], Yiping Tong[1], Dangqun Cui[2] and Aimin Zhang[1,†]

[1] State Key Laboratory of Plant Cell and Chromosome Engineering, National Center for Plant Gene Research, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, 1 West Beichen Road, Chaoyang District, Beijing 100101, China
[2] Department of Agronomy/Key Laboratory of Physiological Ecology and Genetic Improvement of Food Crops in Henan Province, Henan Agricultural University, 63 Nongye Road, Zhengzhou 450002, China

*These authors contributed equally to this work.
† To whom correspondence should be addressed. E-mail: amzhang@genetics.ac.cn

Received 10 October 2012; Revised 6 January 2013; Accepted 18 February 2013

Abstract

Low-molecular-weight glutenin subunits (LMW-GS), encoded by a complex multigene family, play an important role in the processing quality of wheat flour. Although members of this gene family have been identified in several wheat varieties, the allelic variation and composition of LMW-GS genes in common wheat are not well understood. In the present study, using the LMW-GS gene molecular marker system and the full-length gene cloning method, a comprehensive molecular analysis of LMW-GS genes was conducted in a representative population, the micro-core collections (MCC) of Chinese wheat germplasm. Generally, >15 LMW-GS genes were identified from individual MCC accessions, of which 4–6 were located at the *Glu-A3* locus, 3–5 at the *Glu-B3* locus, and eight at the *Glu-D3* locus. LMW-GS genes at the *Glu-A3* locus showed the highest allelic diversity, followed by the *Glu-B3* genes, while the *Glu-D3* genes were extremely conserved among MCC accessions. Expression and sequence analysis showed that 9–13 active LMW-GS genes were present in each accession. Sequence identity analysis showed that all i-type genes present at the *Glu-A3* locus formed a single group, the s-type genes located at *Glu-B3* and *Glu-D3* loci comprised a unique group, while high-diversity m-type genes were classified into four groups and detected in all *Glu-3* loci. These results contribute to the functional analysis of LMW-GS genes and facilitate improvement of bread-making quality by wheat molecular breeding programmes.

Key words: Allele, common wheat, composition, low-molecular-weight glutenin subunits.

Introduction

Common wheat (*Triticum aestivum* L.) is one of the 'big three' cereal crops used for human food (Shewry, 2009) since wheat grains confer their viscoelastic properties to wheat dough (Shewry *et al*., 1995). These viscoelastic properties allow dough to be incorporated into a wide range of daily food products, which are affected by glutenin and gliadin proteins in wheat seeds (Shewry *et al*., 1995; D'Ovidio and Masci, 2004; Juhász and Gianibelli, 2006). Glutenin proteins are composed of two groups of subunits, namely high-molecular-weight and low-molecular-weight glutenin subunits (HMW-GS, 65–90 kDa; LMW-GS, 30–45 kDa) (Payne, 1987; D'Ovidio and Masci, 2004). The LMW-GS account for about one-third of the seed protein and 60% of glutenin proteins, and play an important role in determining dough properties and the quality of wheat food products (Gupta *et al*., 1991, 1994; Branlard *et al*., 2001; Eagles *et al*., 2002; Howitt *et al*.,

2006). Thus, elucidating the composition and variation of LMW-GS genes in common wheat and investigating the relationship between allelic variants and end-use quality are of interest for wheat quality improvement (Gupta *et al*., 1994; He *et al*., 2005; Bekes *et al*., 2006; Juhász and Gianibelli, 2006; Liu *et al*., 2010; Zhang *et al*., 2012).

LMW-GS genes form a multigene family in common wheat, generally located at the *Glu-A3*, *Glu-B3*, and *Glu-D3* loci on the short arms of homoeologous group 1 chromosomes (Jackson *et al*., 1983). The copy number of LMW-GS genes was estimated to range from 10–20 to 30–40 (Ikeda *et al*., 2002; D'Ovidio and Masci, 2004; Juhász and Gianibelli, 2006; Huang and Cloutier, 2008; Dong *et al*., 2010; Zhang *et al*., 2011*a*, *b*). In Norin 61, 12 groups of LMW-GS genes were identified by screening the cDNA library (Ikeda *et al*., 2006). In Glenlea, among the 12 active genes, one was assigned to chromosome 1A, two to chromosome 1B, and nine to chromosome 1D (Huang and Cloutier, 2008). In Xiaoyan 54, 14 unique LMW-GS genes were identified using BAC (bacterial artificial chromosome) library screening and proteomics analysis, of which four were located at *Glu-A3*, three at *Glu-B3*, and seven at *Glu-D3*. Of the 11 active genes, two, two, and seven were i-, s-, and m-type genes, respectively (Dong *et al*., 2010). The above three varieties contained different LMW-GS gene compositions, suggesting that this gene family has high molecular diversity among wheat varieties (Dong *et al*., 2010). Moreover, these LMW-GS proteins had similar physical and chemical properties and molecular weights to the gliadins, which are a type of alcohol-soluble, monomeric seed storage protein. The high copy number and their co-migration with gliadins by SDS–PAGE (sodium dodecyl sulfate polyacrylamide gel electrophoresis) and MALDI-TOF MS (matrix-assisted laser desorption ionization-time of flight mass spectrometry) made separation of LMW-GS proteins and isolation of all LMW-GS genes from a particular wheat variety difficult (Howitt *et al*., 2006). Thus, characterization of the allelic variation of LMW-GS genes in wheat germplasm remains challenging.

To dissect the LMW-GS complex, a nomenclature system was developed based on their relative mobility in SDS-PAGE (Singh *et al*., 1991). Recently, their encoding genes were isolated using PCR with gene-specific primers. One m- and two i-type LMW-GS genes (*GluA3-1*, *GluA3-2*, and *GluA3-3*) were isolated from each *Glu-A3* allele (Wang *et al*., 2010). At the *Glu-B3* locus, four LMW-GS genes (*GluB3-1*, *GluB3-2*, *GluB3-3*, and *GluB3-4*) and their allelic variants were isolated from nine *Glu-B3* alleles (*Glu-B3a*–*Glu-B3i*) (Wang *et al*., 2009). Also, six *Glu-D3* genes were identified from individual wheat varieties containing *Glu-D3a*–*Glu-D3e* (Zhao *et al*., 2006, 2007). Meanwhile, allele-specific markers were developed to discriminate LMW-GS genes and their allelic variants in common wheat (Zhao *et al*., 2007; Appelbee *et al*., 2009; Wang *et al*., 2009, 2010). These markers facilitated identification of the known *Glu-A3* and *Glu-B3* alleles used in breeding programmes (Liu *et al*., 2010). However, *Glu-D3* genes were highly conserved, and allelic identification using PCR markers was difficult (Liu *et al*., 2010). In contrast, these allele-specific primers characterized only one or two genes in individual alleles, and could not be used to determine the exact composition of LMW-GS genes in individual varieties.

To determine the composition of LMW-GS genes in individual wheat varieties, the LMW-GS gene molecular marker system and the full-length gene-cloning method were developed (Zhang *et al*., 2011*a*, *b*), which enabled identification and characterization of the complete sequences of all LMW-GS genes in any wheat variety. In the present study, using both methods, LMW-GS genes were investigated in the micro-core collections (MCC) of Chinese wheat germplasm, which covers >70% of the genetic diversity of Chinese wheat germplasm (Hao *et al*., 2011). The composition, organization, allelic variation, and expression of LMW-GS genes in 262 MCC accessions were comprehensively investigated.

Materials and methods

Wheat germplasm

The MCC of Chinese wheat germplasm were obtained from the Institute of Crop Science, Chinese Academy of Agricultural Sciences (CAAS). This was a representative sample of Chinese wheat diversity. This collection consisted of 262 accessions including 88 modern varieties, 157 landraces, and 17 foreign varieties, which accounted for >70% of the genetic diversity of the national collection.

LMW-GS gene analysis

Genomic DNAs of 262 MCC accessions were extracted from young leaves of seedlings with the cetyltrimethyl ammonium bromide (CTAB method) following Saghai-Maroof *et al*. (1984). The LMW-GS genes in all the MCC accessions were separated using the LMW-GS gene molecular marker system (Zhang *et al*., 2011b). Based on data from the marker system, 45 accessions containing almost all allelic variants were selected for RNA analysis. Total RNA was prepared from developing seeds at 15–21 dpa (day post anthesis) using TRIzol® Reagent, according to the manufacturer's protocol (Invitrogen, Carlsbad, CA, USA). The RNA was converted into cDNA using Moloney murine leukemia virus (M-MLV) reverse transcriptase (Promega, Madison, WI, USA). The expressed LMW-GS genes were detected using the molecular marker system (Zhang *et al*., 2011*b*). To obtain the full-length sequence of these LMW-GS genes, 30 representative varieties containing all the main allelic variants of each gene were selected. All genes were cloned and identified using the full-length gene cloning method (Zhang *et al*., 2011*a*). To clone rare allelic variants, gene-specific primers were developed (Supplementary Table S1 available at *JXB* online). Sequence analysis and characterization were performed using Lasergene software (DNAStar; http://www.dnastar.com/), ClustalW2 (http://www.ebi.ac.uk/Tools/msa/clustalw2/), and MEGA 5 software (Kumar *et al*., 2008).

Results

Composition and variation of LMW-GS genes in MCC accessions

LMW-GS genes in the MCC of Chinese wheat germplasm were amplified using the LMW-GS gene molecular marker system consisting of three independent sets of conserved

primers (Zhang *et al*., 2011*b*). In general terms, at least 15 LMW-GS genes were identified in each wheat variety (Table 1). To characterize these LMW-GS genes further, 30 representative accessions containing the main allelic variants were selected, and all of their LMW-GS genes were cloned and sequenced using the full-length gene cloning method and gene-specific primers (Table 1; Supplementary Table S1 at *JXB* online) (Zhang *et al*., 2011*a*). In total, 466 LMW-GS gene sequences were identified and deposited in GenBank (JX877778–JX878243).

For each LMW-GS gene, several allelic variants were detected from the MCC (Table 1), some of which were identified previously (Ikeda *et al*., 2006; Zhao *et al*., 2006, 2007; Huang and Cloutier, 2008; Wang *et al*., 2009, 2010; Dong *et al*., 2010; Zhang *et al*., 2011*a*, *b*). Using available mapped LMW-GS genes and their allelic relationship with the genes identified from the MCC, all these genes were assigned to specific wheat chromosomes. In individual accessions, 4–6 genes were located at the *Glu-A3* locus, 3–5 at the *Glu-B3* locus, and eight at the *Glu-D3* locus (Table 1). These genes were named according to their DNA fragment size and chromosomal location. For example, the gene corresponding to DNA fragment 441.5, located at the *Glu-D3* locus, was designated *D3-441*. Most of these genes contained several allelic variations across the MCC. To simplify description, the major allelic variant was selected to represent each LMW-GS gene and it was named according to the following scheme: 'DNA fragment+gene' (e.g. the *A3-402* gene), while the allelic variants of each gene were named according to 'DNA fragment+allele' (e.g. the *A3-402* allele).

LMW-GS genes at the Glu-A3 locus At the *Glu-A3* locus, 4–6 LMW-GS genes were detected in each accession and several allelic variants were identified for each gene in the MCC (Table 1; Fig. 1). With regard to the *A3-391* gene, five allelic variants, *A3-353*, *A3-370a*, *A3-370b*, *A3-373*, and *A3-391*, shared >97% identity (Supplementary Fig. S1 at *JXB* online). The *A3-391* allele predominated in 196 accessions, while *A3-353* and *A3-373* were rare variants, each present in only two MCC accessions (Fig. 1a). Sequence analysis of the *A3-391* genes of 30 varieties confirmed that allelic variants showed length polymorphisms in the repetitive regions, and that each variant contained its own single nucleotide polymorphisms (SNPs) (Supplementary Fig. S1). *A3-353*, *A3-373*, and *A3-391* were highly conserved across the MCC population, whereas the *A3-370* allele could be further divided into two variants (*A3-370a* and *A3-370b*) due to SNPs in all available sequences (Supplementary Fig. S1). *A3-353*, *A3-370a*, *A3-370b*, and *A3-391* alleles contained immature stop codons, and only the rare allele *A3-373* possessed an intact open reading frame (ORF) encoding an m-type subunit (Supplementary Table S2). Thus, the *A3-391* gene was universal in common wheat, even though only five sequences with >98% identities were deposited in GenBank.

The *A3-400* gene was also common in wheat varieties. Seven allelic variants with different repetitive region lengths, *A3-374*, *A3-388*, *A3-394*, *A3-400*, *A3-402*, *A3-408*, and *A3-411*, were identified from MCC accessions (Fig. 1a). Sequence alignments suggested that the *A3-374*, *A3-388*, *A3-400*, *A3-408*, and *A3-411* alleles were conserved among wheat varieties, whereas both *A3-394* and *A3-402* comprised two variants, *A3-394a* and *A3-394b*, and *A3-402a* and *A3-402b*, respectively, due to indels and SNPs in the available sequences (Supplementary Fig. S2 at *JXB* online). Sequence analysis also demonstrated that *A3-402a* and *A3-400* shared the same SNP, containing a premature stop codon, while all remaining allelic variants (i.e. *A3-374*, *A3-388*, *A3-394a*, *A3-394b*, *A3-402b*, *A3-408*, and *A3-411*) contained intact coding sequences that might encode m-type LMW-GS in common wheat (Supplementary Fig. S2).

Except for the m-type genes above, all the others identified at the *Glu-A3* locus were i-type genes (Fig. 1a; Supplementary Table S2 at *JXB* online). The coding sequences of alleles *A3-480*, *A3-484*, *A3-487*, *A3-502*, and *A3-508* contained a specific length of repetitive regions and unique SNPs. For the major variant *A3-502*, eight allelic variants (*A3-502a–A3-502h*) were recognized due to SNPs and indels in the coding sequences (Fig. 1a; Supplementary Fig. S3). For the other i-type genes, eight genes/haplotypes, *A3-620*, *A3-626*, *A3-643*, *A3-646*, *A3-649*, *A3-573*/*A3-640*, *A3-567*/*A3-590*, and *A3-565*/*A3-568*/*A3-662*, were detected, of which alleles *A3-626* and *A3-646* were further divided into two allelic variants, respectively, due to SNPs in the coding sequences. Moreover, two genes, *A3-649-1* and *A3-649-2* with unique SNPs and indels, which shared a 649 bp DNA fragment, were identified from individual accessions; the *A3-567-1* and *A3-567-2* genes exhibited similar commonality. After analysing the composition of i-type genes in 262 MCC accessions, it was found that all *A3-502* variants were tightly linked with other unique i-type genes. The *A3-502a* and *A3-502b* alleles were coupled with *A3-620*, *A3-502c* with *A3-626a*, *A3-502d* with *A3-643*, *A3-502e* with *A3-646a*, *A3-502f* with *A3-646b*, *A3-502g* with *A3-573* and *A3-640*, and *A3-502h* with *A3-649-1* and *A3-649-2*. In total, 12 i-type haplotypes were identified in MCC accessions (Fig. 1a).

LMW-GS genes at the Glu-B3 locus Generally, excluding 11 1BL/1RS translocation lines, 3–5 *Glu-B3* genes were identified in each MCC accession, and *B3-530* and *B3-548* genes were universal (Fig. 1b). With regard to the *B3-530* gene, three allelic variants, *B3-510*, *B3-516*, and *B3-530*, were identified (Table 1; Fig. 1b), and shared high sequence identity (>99%). The *B3-530* allele was further divided into three variants (*B3-530a*, *B3-530b*, and *B3-530c*) due to the unique SNPs (Supplementary Fig. S4 at *JXB* online). All allelic variants of the *B3-530* gene possessed intact ORFs and their deduced proteins belonged to m-type LMW-GS. For the *B3-548* gene, the *B3-548* allele predominated in 241 accessions, while the rare allelic variant *B3-557* was detected in only one (Table 1; Fig. 1b). However, both variants contained premature stop codons in their coding sequences, suggesting them to be pseudogenes.

The other *Glu-B3* genes were identified only in partial MCC accessions (Fig. 1). The *B3-570* gene was identified in 44 MCC accessions, and its intact ORF contained 344 amino acid residues and encoded an m-type gene with the novel N-terminal sequence, METSQIPGLEKPS. The *B3-578*, *B3-621*, and *B3-544*

Table 1. LMW-GS genes and their allelic variants identified from MCC accessions using the LMW-GS gene molecular marker system.

Genes and allelic variants[a]				LMWGS1[b]				LMWGS2[b]				LMWGS3[b]			
A3-391	*A3-353*	*A3-370*	*A3-373*	391.8	353.1	370.8	373.2	484.8	460.4	477.5	480.1	375.7	350.2	368.8	371.2
A3-400	*A3-374*	*A3-388*	*A3-394*	400.1	374.5	388.1	N[c]	506.3	481.4	494.7	N	399.1	N	387.2	N
	A3-402	*A3-408*	*A3-411*		402.8	408.3	411.1		509.5	514.7	517.6		402.5	408.5	411.4
A3-502	*A3-480*	*A3-484*	*A3-487*	N	N	484.3	N	N	N	590.2	N	538.1	513.8	N	522.7
	A3-508				N				N				543.8		
A3-565				565.3				671.4				606.7			
A3-568	*A3-567*			568	567.2			674.1	673.2			609.6	607.2		
A3-620	*A3-590*	*A3-626*	*A3-640*	620.1	589.5	626.2	N/640.6	725.3	693.7	730.5	N/745.4	666.6	632.3	673.3	619.4/688.5
	A3-643	*A3-646*	*A3-649*		643.2	646	649.5		N	750.6	753.6		691	695.6	698.6
	A3-662				662.6				766.8				715.8		
B3-530	*B3-510*	*B3-516*		530.8	510.5	516		636.9	616.6	622.4		538.2	516.7	522.4	
B3-548	*B3-557*			548.8	557.6			654.8	663.7			556.5	565.7		
B3-570				570.8				676.2				N			
B3-578				577.7				683.4				587.8			
B3-544	*B3-590*	*B3-593*	*B3-596*	544.2	590.4	593.4	596.2	650.2	696.1	699	701.6	554.7	605.6	608.7	611.8
	B3-598	*B3-601*	*B3-604*		598.9	601.7	604.6		704.5	706.8	709.6		615.1	618.4	621.5
	B3-607	*B3-610*			607.4	610.2			712.4	715.6			625	628.1	
B3-688	*B3-621*	*B3-624*		688.2	621.1	624.5		791.9	726.1	729.1		707.5	635.3	638.5	
B3-691	*B3-813*			691.3	813.8			794.6	916.3			710.4	855.6		
D3-385				385.7				492.3				383.6			
D3-393	*D3-385'*			393.1	N			499.5	N			390.7	N		
D3-394	*D3-397*			394.6	397.3			500.7	503.3			393.5	395.8		
D3-441	*D3-432*	*D3-444*		441.5	432.7	444.4		547.5	538.7	550.3		441.8	432.6	444.6	
D3-525	*D3-522*	*D3-528*		525.4	522.4	528.2		631.8	628.9	634.5		532.8	529.3	535.9	
D3-575				575				681				584			
D3-578				577.7				683.4				587.8			
D3-586	*D3-583*	*D3-589*	D3-591	586.5	583.6	589	591.8	691.3	688.5	694	696.8	597.9	594.6	601	604.3
	D3-594	*D3-597*			594.8	597.5			699.6	702.6			607.6	610.5	

[a] LMW-GS genes and allelic variants were named according to the size of the DNA fragments amplified using the primer LMWGS1. For each gene, the major allelic variant was designated as the LMW-GS gene and the remainder as its variants.
[b] Three sets of conserved primers in the LMW-GS gene molecular marker system (Zhang *et al*., 2011*b*).
[c] Not detected with the specific primers.

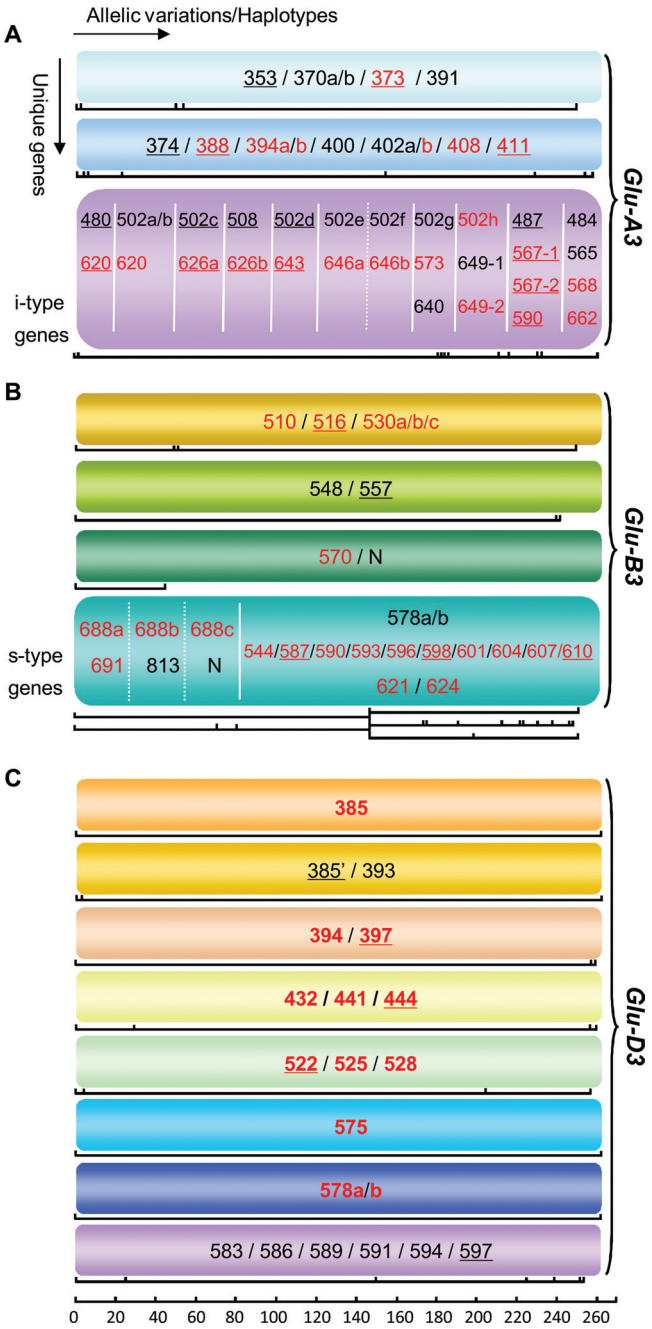

Fig. 1. Composition of LMW-GS genes at the *Glu-3* loci in micro-core collections (MCC) of Chinese wheat germplasm. The diagrams illustrate the LMW-GS genes and their allelic variants at the *Glu-A3*, *B3*, and *D3* loci identified from MCC accessions. The horizontal axis of each diagram shows the allelic variants of individual genes or haplotypes identified from the MCC. The vertical axis displays the composition of unique genes and haplotypes in individual accessions. The length of line segments represents the number of accessions containing the corresponding allelic variants. Underlined allelic variants were rare in the MCC, and allelic variants in red were active in common wheat. The allelic variants indicated by numbers and letters, such as 370a and 370b, shared the same DNA fragment but had different nucleotide sequences. (A) The compositions of genes and their allelic variants at the *Glu-A3* locus. The i-type genes showed high diversity and were tightly linked, forming specific haplotypes. (B) The compositions of genes and their allelic variants at the *Glu-B3* locus. The s-type genes were tightly linked and formed specific haplotypes. (C) The composition of eight genes and their allelic variants at the *Glu-D3* locus.

genes were tightly linked at the *Glu-B3* locus and formed several haplotypes (Fig. 1b). The *B3-578* gene was classified into two variants (*B3-578a* and *B3-578b*) based on two SNPs. The *B3-544* genes were conserved among accessions, and the allelic variants *B3-544* and *B3-587*–*B3-607* shared >99% sequence identity (Supplementary Fig. S5 at *JXB* online). Also, *B3-621* and *B3-624* differed in only a CAA indel and an SNP (Supplementary Fig. S6). Moreover, *B3-578* contained two premature stop codons in its coding sequences, while the other genes possessed intact ORFs, which might be active in common wheat. Sequence analysis showed that all three genes were s-type, the deduced protein sequences of which contained a MENSHIPGLERPS peptide at the N-terminus. *B3-688* could be divided into three groups (*B3-688a*, *B3-688b*, and *B3-688c*) in MCC accessions, although several irregular SNPs were found in the coding sequences (Supplementary Fig. S6). These irregular SNPs were tightly linked with *B3-691* or *B3-813* at the *Glu-B3* locus, forming different haplotypes, namely *B3-688a*/*B3-691*, *B3-688b*/*B3-813*, and *B3-688c*/*N* (Fig. 1b). *B3-688a* was coupled with *B3-691*, and shared >99% identity, the only differences being a CAA indel and an SNP. *B3-813* was identified for the first time in common wheat and contained a premature stop codon in the ORF, while both *B3-688* and *B3-691* had unbroken ORFs encoding s-type LMW-GS in common wheat.

LMW-GS genes at the Glu-D3 locus Eight LMW-GS genes (*D3-385*, *D3-393*, *D3-394*, *D3-441*, *D3-525*, *D3-575*, *D3-578*, and *D3-586*) were detected at the *Glu-D3* locus in individual wheat varieties (Table 1; Fig. 1c). Neither *D3-385* nor *D3-575* preserved any allelic variants and were universal in all MCC accessions. In terms of the *D3-393* gene, the *D3-393* allele was present in 259 MCC accessions, while the other rare allelic variant, *D3-385'*, was found in only three (Fengkang 2, Guinong 10, and Lovrin 10). For the *D3-394* gene, both allelic variants, *D3-394* and *D3-397*, shared >99% sequence identities, the difference being an indel (CAA) and two SNPs; the latter allele was detected in only a single landrace, Yizhimai. Three conserved allelic variants of *D3-441* (*D3-432*, *D3-441*, and *D3-444*) were identified with CAA indels in the repetitive region (Supplementary Fig. S7 at *JXB* online). The three allelic variants of the *D3-525* gene, namely *D3-522*, *D3-525*, and *D3-528*, exhibited a similar phenomenon (Supplementary Fig. S8). For the *D3-578* gene, *D3-578a* was identical in length but had a different nucleotide sequence from that of the the *D3-578b* variant (Supplementary Fig. S9). This gene encoded the only s-type LMW-GS at the *Glu-D3* locus in common wheat. The nucleotide sequences of *D3-586* and its allelic variants were identical, except for the CAA indels in the repetitive region that contributed to the length polymorphism (Supplementary Fig. S10). Among the eight LMW-GS genes at the *Glu-D3* locus, *D3-393* was a

pseudogene due to a frameshift mutation and *D3-586* was a pseudogene due to a premature stop codon. The remaining six genes contained intact ORFs that encoded one s-type (*D3-578*) and five m-type LMW-GS at the *Glu-D3* locus, which contained the largest number of active genes of all the *Glu-3* loci.

Organization of LMW-GS genes in MCC accessions

Using the LMW-GS gene molecular marker system and the full-length gene cloning method (Zhang *et al.*, 2011*a*, *b*), almost all LMW-GS genes were detected in the MCC and were cloned and characterized in 30 MCC accessions, which facilitated investigation of the organization of LMW-GS genes and their linkage relationship.

At the *Glu-A3* locus, 4–6 genes were generally isolated from individual MCC accessions, including *A3-391*, *A3-400*, and 2–4 i-type genes (e.g. *A3-502a* and *A3-620*; *A3-502g*, *A3-573*, and *A3-640*; and *A3-484*, *A3-565*, *A3-568*, and *A3-662*; Figs 1a, 2a). Although several allelic variants were identified for each gene, only 11 main types of *Glu-A3* genotypes were detected in MCC accessions, suggesting that these genes were tightly linked (Fig. 2a). Two genotypes containing *A3-391*, *A3-400* or *A3-402a*, *A3-502*, and *A3-620* accounted for ~70% of the MCC accessions, while each of the remaining genotypes accounted for <7% (Fig. 2a). Moreover, concerning the distribution of different genotypes in foreign accessions, Chinese modern varieties, and landraces, genotypes containing the *A3-649-1*/*-2* genes were detected only in landraces, while those containing *A3-394a*/*b* were present in both foreign accessions and Chinese modern varieties (Fig. 2a).

Based on the genotype data of MCC accessions, the linkage relationship among LMW-GS genes at the *Glu-A3* locus was analysed (Fig. 2a). First, four main haplotypes of i-type genes were detected in MCC accessions, namely *A3-502a*/*b*/*A3-620*, *A3-502*/*A3-646*, *A3-502h*/*A3-649-1*/*A3-649-2*, and *A3-484*/*A3-565*/*A3-568*/*A3-662* (Figs 1a, 2a). The i-type genes were completely linked and formed specific haplotypes in common wheat. Secondly, the *A3-391* allele was coupled with alleles *A3-400* and *A3-402a*, and the *A3-370a*/*b* alleles co-segregated with *A3-394b*, *A3-402b*, *A3-408*, and *A3-411* (Fig. 2a). m-Type genes may have been tightly linked with each other at the *Glu-A3* locus. Thirdly, the *A3-370a*/*b* alleles were generally coupled with *A3-502*(*e*/*f*)/*A3-646* and *A3-484*/*A3-565*/*A3-568*/*A3-662*, while the *A3-391* allele was linked with *A3-502*(*a*/*b*)/*A3-620*, *A3-502h*/*A3-649-1*/*A3-649-2*, and *A3-502g*/*A3-573*/*A3-640* (Fig. 2a). Thus, the *A3-391* gene generally linked with i-type genes in MCC accessions.

At the *Glu-B3* locus, although 3–5 genes were present in individual accessions, their allelic variants consisted of 12 main genotypes in the MCC accessions (Figs 1b, 2b). Genotypes containing allelic variants *B3-621*, *B3-624*, or *B3-688* covered all MCC accessions, excluding 11 1BL/1RS translocation lines, and genotypes containing the haplotypes *B3-688*/*N* or *B3-688*/*B3-691* accounted for 54% of the MCC (Fig. 2b). Tight linkage of LMW-GS genes at the *Glu-B3* locus was also observed. The *B3-510* allele was tightly coupled with the *B3-570* gene, forming the haplotype *B3-510*/*B3-570*, which was generally linked with another haplotype *B3-688*/*B3-691* (Fig. 2b). In contrast, s-type genes consisted of two groups of haplotypes at the *Glu-B3* locus; one contained the *B3-621* gene and the other the *B3-688* gene (Fig. 1b). The former group formed various haplotypes, which contained three unique LMW-GS genes, namely *B3-578*, *B3-544* and *B3-621*, as well as their allelic variants. Alleles *B3-544*, *B3-601*, *B3-604*, and *B3-607* regularly co-segregated with *B3-621*, while the other variants *B3-590*, *B3-593*, and *B3-596* were usually coupled with *B3-624* (Fig. 2b). In terms of the distribution of these genes in MCC, 1BL/1RS lines and genotypes containing *B3-601* and *B3-604* were identified only in foreign or modern varieties, whereas the genotypes containing *B3-624* and the genotype *B3-530*/*B3-548*/*B3-688*/*B3-691* were detected mostly in landraces (Fig. 2b).

At the *Glu-D3* locus, although eight LMW-GS genes were identified from individual accessions, few haplotypes were characterized throughout the entire MCC population because of the high conservation of LMW-GS genes at this locus (Fig. 2c). Among them, *D3-578a* was linked with the *D3-432* allele, whereas *D3-578b* was generally coupled with *D3-441* (Fig. 2c). The *D3-586* gene showed high length polymorphisms among MCC accessions for six allelic variants *D3-583*/*586*/*589*/*591*/*594*/*597*, which resulted in detection of 14 main genotypes at the *Glu-D3* locus in MCC accessions (Fig. 2c). Two genotypes, *D3-385*/*D3-393*/*D3-394*/*D3-441*/*D3-525*/*D3-575*/*D3-578*/*D3-586* and *D3-385*/*D3-393*/*D3-394*/*D3-441*/*D3-525*/*D3-575*/*D3-578*/*D3-589*, differed only in the pseudogene *D3-586*, which accounted for 64.8% of the MCC accessions (Fig. 2c).

Expression of LMW-GS genes in the MCC accessions

Since only the active genes in this family affected bread-making quality, the mRNA of the developing seeds was investigated (Fig. 3) and active LMW-GS genes were identified by comparing the mRNA and genomic DNA data (Fig. 3).

At the *Glu-A3* locus, among the 4–6 genes detected in genomic DNA (Fig. 2a), the *A3-370a*/*b* and *A3-391* alleles were not detected in mRNA of developing seeds (Fig. 3), while their allelic variant *A3-373* may have been active in the intact ORF. In terms of the *A3-400* gene, the *A3-400* and *A3-402a* alleles were inactive, whereas both *A3-402b* and *A3-408* were expressed during seed filling (Fig. 3). Also, the rare allelic variants *A3-374*, *A3-388*, *A3-394a*/*b*, and *A3-411* may encode LMW-GS proteins in their intact ORFs. None of the allelic variants of the *A3-502* gene was expressed due to premature stop codons, except for the active allele *A3-502h*. Among the other i-type genes, *A3-573*, *A3-620*, *A3-646*, *A3-649-2*, *A3-568*, and *A3-662* were detected in developing seeds (Fig. 3), and the rare allelic variants *A3-626a*/*b*, *A3-643*, *A3-567-1*, *A3-567-2*, and *A3-590* contained intact ORFs, which might also be active in common wheat. Thus, in a particular wheat variety, 1–3 LMW-GS genes might be active at the *Glu-A3* locus (Fig. 2a). For example, accessions with the genotype *A3-391*/*A3-400*/*A3-502*/*A3-620* contained only one active

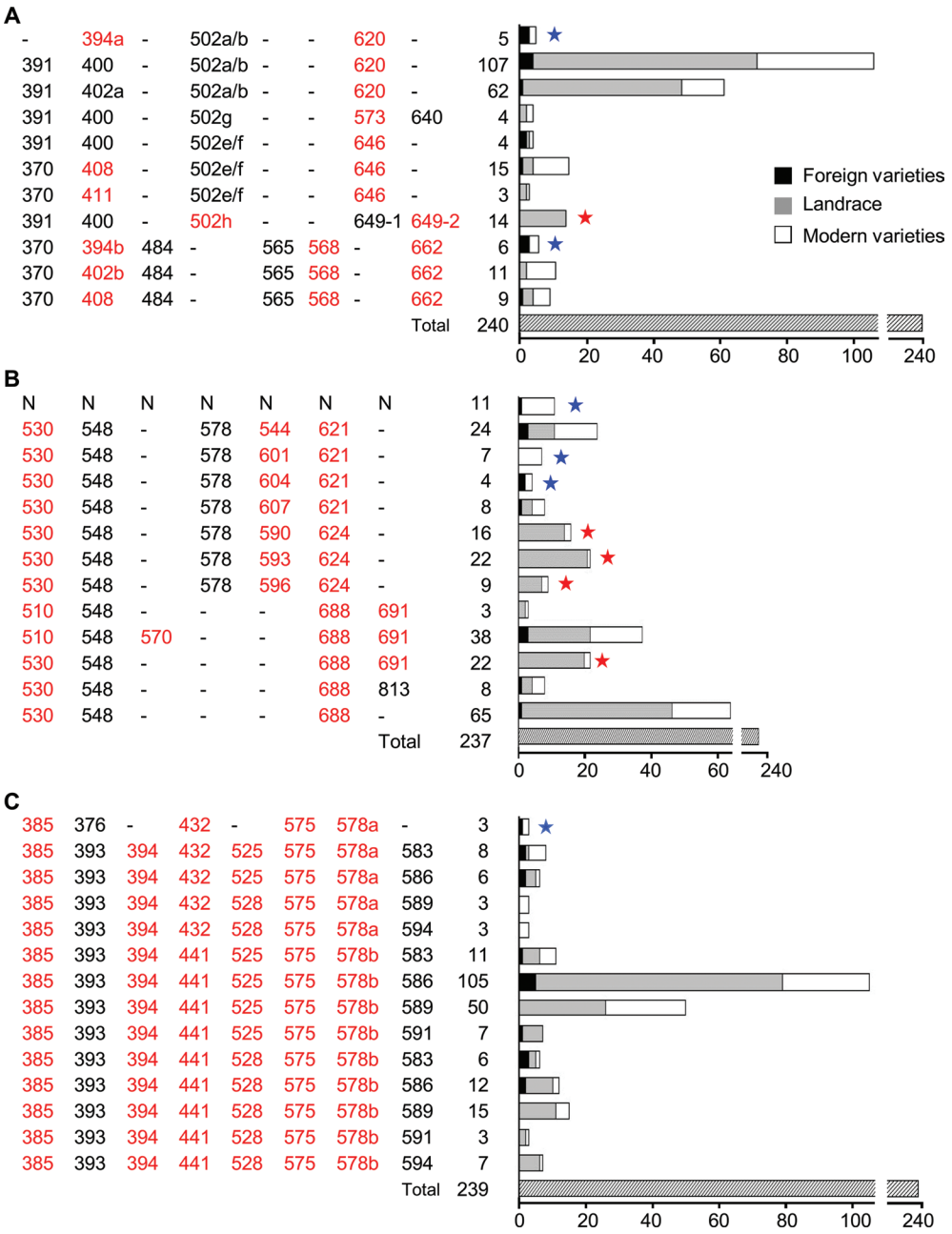

Fig. 2. Main genotypes of LMW-GS genes at *Glu-3* loci identified from the MCC. Genotypes present in more than three MCC accessions are shown. LMW-GS genes at the same locus were generally linked and formed limited types of genotypes. Genes in red were active in common wheat. Moreover, the modern varieties, landraces, and foreign varieties are indicated by different colours. The genotypes most common in landraces are indicated by red asterisks, and those found only in modern and foreign varieties by blue asterisks. (A) Eleven genotypes at the *Glu-A3* locus. (B) Thirteen genotypes at the *Glu-B3* locus. Of these, 11 accessions without LMW-GS genes belong to 1B/1R translocation lines. (C) Fourteen genotypes at the *Glu-D3* locus.

LMW-GS gene, *A3-620*, while those with the i-type haplotype *A3-484*/*A3-565*/*A3-568*/*A3-662* possessed three active genes at the *Glu-A3* locus (Fig. 2a).

At the *Glu-B3* locus, 3–5 genes were detected in the genomic DNA of individual accessions (Fig. 2b). Based on the electropherogram of the LMW-GS gene marker system, the *B3-548*, *B3-578a*/*b*, and *B3-813* genes were not detected in developing seeds (Fig. 3), while the remaining m-type genes (*B3-530* and the newly identified *B3-570*) and s-type genes (*B3-544*, *B3-621*, and *B3-688*) were generally expressed in wheat varieties, although *B3-570* was only present in partial MCC accessions (Figs 2b, 3). With regard to active genes of different haplotypes, the number of active *Glu-B3* genes varied from two to four in one wheat variety, including one or two m-type-encoding genes and one or two s-type-encoding genes (Fig. 2b).

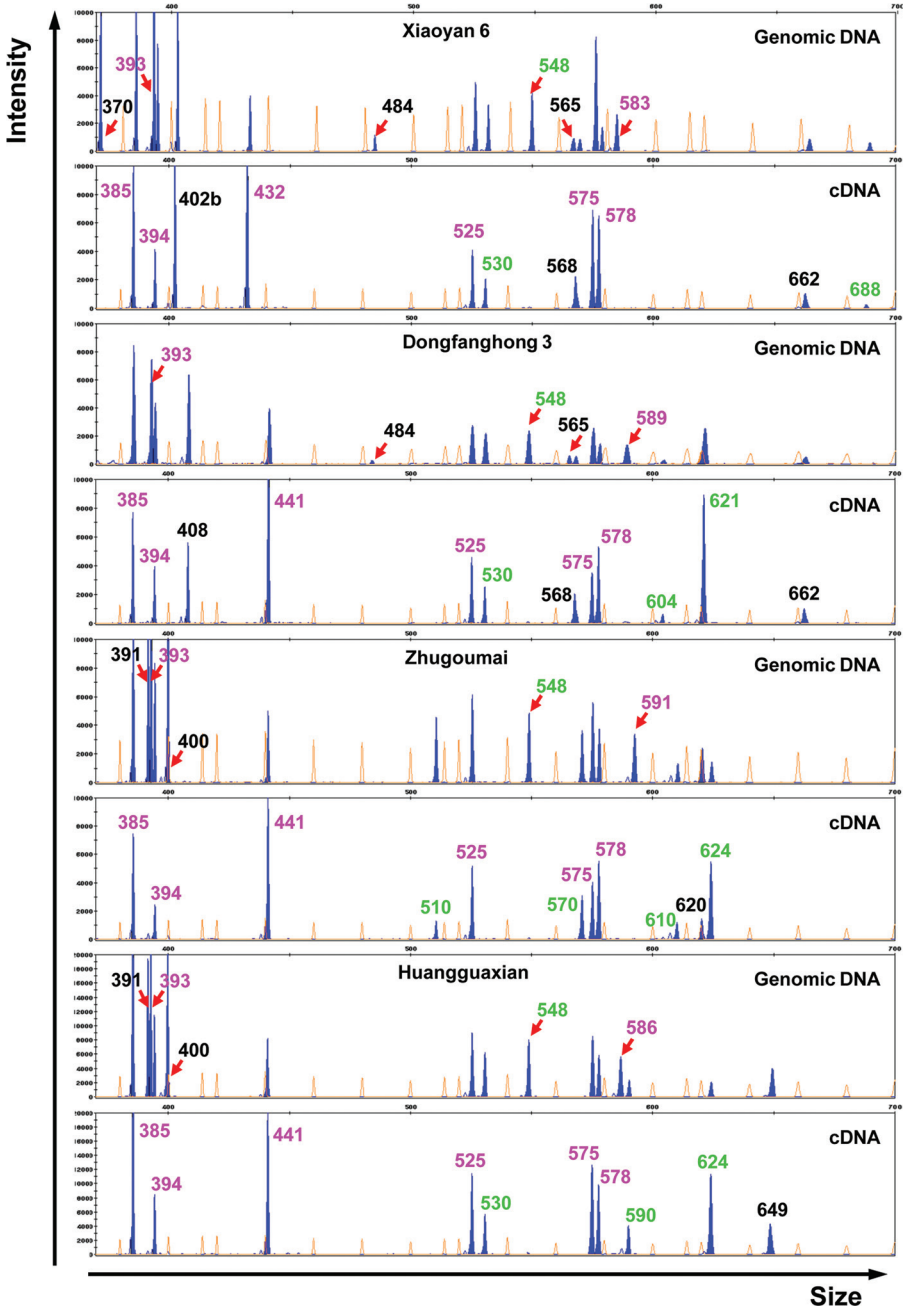

Fig. 3. Expression analysis of LMW-GS genes in MCC accessions. Electropherograms show the patterns of DNA fragments detected in genomic DNA and cDNA of individual accessions using the LMWGS1 primers. The results of four of 45 representative accessions are shown. The horizontal axis shows the size of the detected DNA fragments, while the vertical axis displays the signal intensity (i.e. the concentration of DNA fragments in the PCR products). The orange peaks indicate the GeneScan 1200 LIZ size standard fragments, while the blue peaks represent the DNA fragments in the PCR products. In the genomic DNA electropherogram, peaks indicated by arrows and numbers were detected only in genomic DNA, all of which were pseudogenes. Peaks in the cDNA electropherograms indicated by numbers correspond to active LMW-GS genes. *Glu-A3* genes are shown in black, *Glu-B3* genes in green, and *Glu-D3* genes in pink.

At the *Glu-D3* locus, of the eight LMW-GS genes detected in genomic DNA, *D3-385*, *D3-394*, *D3-441*, *D3-525*, *D3-575*, and *D3-578* were detected in developing seeds of MCC accessions (Fig. 3), and the rare allelic variants *D3-397*, *D3-444*, and *D3-522* might be active in their intact ORF. The remaining two genes, *D3-393* and *D3-586*, were not identified in the developing seeds probably due to premature stop codons (Fig. 3). Thus, one s-type (*D3-578*) and five m-type genes (*D3-385*, *D3-394*, *D3-441*, *D3-525*, and *D3-575*) were generally expressed at the *Glu-D3* locus in individual wheat varieties (Fig. 2c). These expression analyses revealed that individual accessions with different allelic variants or haplotypes might possess different numbers of active genes. Generally, none or one m-type and one or two i-type active genes were detected

at the *Glu-A3* locus, and one or two m-type and one or two s-type genes at the *Glu-B3* locus were expressed, whereas one s-type and five m-type genes comprised the active genes at the *Glu-D3* locus. Thus, the number of active genes in individual MCC accessions varied from nine to 13.

Characteristics of LMW-GS genes identified from MCC accessions

All proteins encoded by the genes identified in the present study were typical of LMW-GS, which had similar structures to previously characterized LMW-GS (D'Ovidio and Masci, 2004; Juhász and Gianibelli, 2006). Each deduced protein contained four main structural domains: a signal peptide, a short conserved N-terminal domain, a repetitive domain, and a C-terminal domain, except for i-type proteins, which lacked the N-terminal domain (Supplementary Fig. S11 at *JXB* online). Based on the N-terminal sequence of mature proteins, three types of LMW-GS (m-, s-, and i-types) were recognized. The m-type proteins were the most abundant in all genotypes analysed, and their molecular mass varied from 31.8 kDa (D3-385) to 39.6 kDa (D3-575). The s-type proteins generally had a higher molecular mass than did m-type subunits, which ranged from 37.0 kDa (B3-544) to 42.5 kDa (B3-691). Also, the i-type proteins had higher molecular weights (39.2–43.0 kDa), despite lacking the N-terminal sequences.

Cysteine residues played a vital role in determining the structural and functional characteristics of wheat proteins (Shewry *et al*., 1995; D'Ovidio and Masci, 2004). All deduced proteins identified in this study possessed eight cysteine residues, except the putative amino acid sequences from pseudogenes *A3-502d* and *D3-385'*, which contained seven and nine cysteine residues, respectively. However, both pseudogenes do not play a role in glutenin polymers and bread-making quality. The locations of the first (or third for i-type genes) and seventh cysteines were highly diverse, while the remaining six cysteines were conserved among all LMW-GS genes (Supplementary Fig. S11 at *JXB* online). Based on the relative locations of cysteines, LMW-GS proteins were divided into six groups (Supplementary Fig. S11).

All LMW-GS genes and their allelic variants were subjected to cluster analysis using ClustalW2 and MEGA 5. Also, the six main groups were further divided, which was consistent with the grouping data based on the cysteine positions of the deduced proteins (Fig. 4). The i-type genes located at the *Glu-A3* locus formed a single group (i_A), and all the s-type genes at the *Glu-B3* locus and *Glu-D3* locus were located in a single branch (s_{BD}) (Fig. 4). All the remaining LMW-GS genes were m-type, which were further divided into four groups (Fig. 4). Variants of both the *D3-441* and *D3-525* genes formed single groups (m_D-2 and m_D-1, respectively), which were unique at the *Glu-D3* locus. The m_{BD} group was composed of three genes (*B3-530*, *B3-548*, and *B3-570*) from the *Glu-B3* locus and two (*D3-575* and *D3-586*) from the *Glu-D3* locus. In contrast, the m_{AD} group contained five m-type genes, two of which were located at the *Glu-A3* locus and three at the *Glu-D3* locus. Collectively, genes from the *Glu-A3* locus contained all the i-type genes (i_A) and m-type genes (m_{AD}), genes from

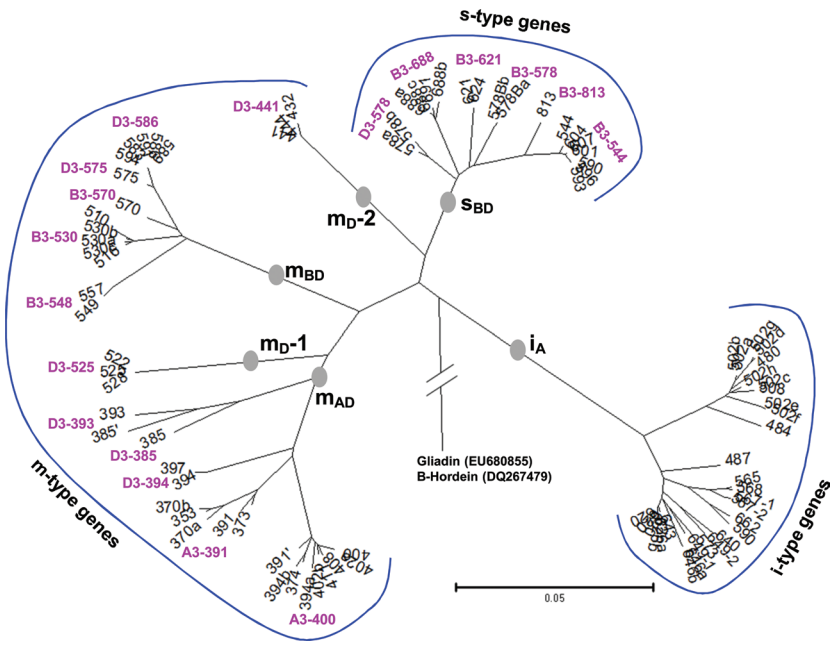

Fig. 4. Phylogenetic reconstruction of all LMW-GS genes and their allelic variants identified from MCC accessions. The phylogenetic tree of LMW-GS genes was constructed using MEGA 5 (Kumar *et al*. 2008). All LMW-GS genes were divided into six groups. The i-type genes located at *Glu-A3* were special and formed a single group (i_A). The s-type genes at *Glu-B3* and *Glu-D3* shared high identity and were located in a single branch (s_{BD}). The other LMW-GS genes were of the m-type, and were divided into four groups (m_{AD}, m_{D-1}, m_{BD}, and m_{D-2}). LMW-GS genes at the *Glu-D3* locus were assigned to five groups, and showed a higher diversity than those at the *Glu-A3* and *Glu-B3* loci. (This figure is available in colour at *JXB* online.)

the *Glu-B3* locus were associated with two groups, including s-type (s_{BD}) and m-type (m_{BD}) genes, and genes from the *Glu-D3* locus were distributed into five of the six groups and showed higher diversity than those at the *Glu-A3* and *Glu-B3* loci (Fig. 4).

Of the LMW-GS genes, i-type genes were the most complex and 12 haplotypes were detected, each of which contained unique LMW-GS genes. Sequence alignments and phylogenetic analysis of all i-type genes demonstrated that the *A3-502* gene was conserved (>85% diversity) among the haplotypes, whereas the other genes could be divided into five subgroups (Supplementary Fig. S12 at *JXB* online). *A3-626* and *A3-643* shared high identity (>98%) and formed the subgroup i_A-1 with the *A3-502* gene. The haplotypes *A3-502e/f* and *A3-646* were named i_A-2. The *A3-502g*/*A3-573*/*A3-640* and *A3-502h*/*A3-649-1*/*A3-649-2* haplotypes contained three i-type genes and were classified into subgroups i_A-3 and i_A-4, respectively. *A3-484*/*A3-565*/*A3-568*/*A3-662* and *A3-487*/*A3-567-1*/*A3-567-2*/*A3-590* represented a unique i-type genotype (i_A-5) in common wheat (Supplementary Fig. S12).

Discussion

LMW-GS genes in common wheat are complex, and their exact copy number remains unclear (Cassidy *et al*., 1998; Ikeda *et al*., 2002; Juhász and Gianibelli, 2006; Huang and Cloutier, 2008; Dong *et al*., 2010). Recently, using BAC library screening, 14 and 19 genes were isolated from the common wheat varieties Xiaoyan 54 and Glenlea, respectively (Huang and Cloutier, 2008; Dong *et al*., 2010). Meanwhile, LMW-GS genes at the *Glu-A3*, *Glu-B3*, and *Glu-D3* loci were identified using gene-specific primers, which suggested that at least 12 genes are present in the common wheat genome (Zhao *et al*., 2006, 2007; Wang *et al*., 2009, 2010). Based on the conserved and polymorphic structures of these genes, the LMW-GS gene marker system and the full-length gene cloning method were developed, which can identify >15 members of this gene family in common wheat (Zhang *et al*., 2011*a*, *b*). In the present study, both methods were used to investigate the MCC of Chinese wheat germplasm, and the complex LMW-GS gene family in common wheat was successfully dissected.

Dissection of LMW-GS genes at individual Glu-3 loci

Glu-A3 locus Two m-type genes and 2–4 i-type LMW-GS genes were generally identified at the *Glu-A3* locus, which was the highest number reported for this locus in individual wheat varieties (Figs 1a, 2a). The m-type gene, *A3-391*, and its allelic variants shared high identities (>99%) with a few sequences in GenBank derived from *T. macha*, *T. durum*, and *T. timopheevii* (Supplementary Table S2 at *JXB* online), which suggests that *A3-391* might be widely present in *Triticum*. The other m-type gene at the *Glu-A3* locus, *A3-400*, was reported by several groups, corresponding to *GluA3-2* genes from Aroona near-isogenic lines (NILs) (Wang *et al*., 2010), the group 6 type IV gene from Norin 61 (Ikeda *et al*., 2002), and *A3-1* from Xiaoyan 54 (Supplementary Table S2) (Dong *et al*., 2010). The present results provided direct evidence for the presence of m-type genes at the *Glu-A3* locus in common wheat, and this gene showed high diversity among MCC accessions with several novel allelic variants (i.e. *A3-374*, *A3-388*, *A3-394a*, and *A3-411*; Fig. 1a). Moreover, the new allelic variants, *A3-388*, *A3-394a*/*b*, *A3-408*, and *A3-411*, contained intact ORFs and may make specific contributions to wheat bread-making quality.

In the present study, using conserved primers, 2–4 i-type genes were identified in individual wheat varieties. (Supplementary Table S2 at *JXB* online). In previous studies, 1–3 i-type genes in only a few wheat varieties were characterized (Supplementary Table S2) (Zhang *et al*., 2004; Ikeda *et al*., 2006; Huang and Cloutier, 2008; Dong *et al*., 2010), which made it difficult to analyse the relationships among haplotypes of these genes. Here, the MCC of Chinese wheat germplasm were investigated in terms of LMW-GS gene composition. The i-type genes were present in the wheat genome as haplotypes rather than single genes, and 12 haplotypes of i-type genes were detected in the MCC. Nucleotide sequence comparisons showed that genes in six of 12 haplotypes identified in this study were similar (>99%) to those isolated from seven *Glu-A3* alleles, for example haplotype *A3-502d*/*643* corresponding to *GluA3-32*/*Glu-A3-12* from *Glu-A3b* (Supplementary Table S2) (Wang *et al*., 2010; Zhang *et al*., 2012). In addition, haplotype *A3-484*/*A3-565*/*A3-568*/*A3-662* contained the three i-type genes identified in Norin 61 and Xiaoyan 54 (*A3-2*, *A3-3*, and *A3-4*), and the haplotype *A3-502f*/*646b* covered the i-type gene detected in Glenlea (EU189087) (Huang and Cloutier, 2008; Dong *et al*., 2010). This confirms that i-type genes in common wheat exist as haplotypes at the *Glu-A3* locus and exhibit high genetic diversity. The identification and characterization of these haplotypes will facilitate the functional analysis of i-type genes and the selection of specific genes using haplotype-specific markers.

Glu-B3 locus Three to five *Glu-B3* genes were detected in individual varieties, of which 1–3 were s-type and two or three were m-type (Figs 1b, 2b). The m-type gene *B3-530* shared >99% sequence identity with *GluB3-4* genes from Aroona and its near-isogenic lines, the *B3-1* gene from Xiaoyan 54 and Jing 411, *1557N24-M* from Glenlea, and the group 2 type I gene from Norin 61 (Supplementary Table S2 at *JXB* online) (Ikeda *et al*., 2002; Huang and Cloutier, 2008; Wang *et al*., 2009; Dong *et al*., 2010), while *B3-548* has been reported only rarely since it is a pseudogene. The third m-type gene, *B3-570*, was newly identified from wheat varieties and was detected in partial MCC accessions containing *B3-510* (Fig. 2b). Thus, at least two m-type genes were present at the *Glu-B3* locus, rather than the one reported previously (Ikeda *et al*., 2002; Huang and Cloutier, 2008; Dong *et al*., 2010). The other genes at the *Glu-B3* locus were of s-type, and were divided into two subgroups based on their gene composition; one containing the *B3-688* gene and the other containing *B3-621* (Figs 1b, 2b). The *B3-688* subgroup of s-type haplotypes corresponded to *B3-2* from Xiaoyan 54 as well as Jing 411 and *GluB3-3* from Aroona-*Glu-B3c*, *B3d*,

B3h, and *B3i* (Supplementary Table S2) (Wang *et al*., 2009; Dong *et al*., 2010; Zhang *et al*., 2012). The other subgroup contained two active genes, *B3-544* and *B3-621*, and one pseudogene, *B3-578*. *B3-544* had 99% sequence identity with *GluB3-1* from Aroona-*Glu-B3a*, *B3b*, *B3f*, and *B3g*. Also, *B3-621* genes were present in Aroona-*Glu-B3a*, *B3b*, *B3f*, and *B3g*, corresponding to *GluB3-2* (Wang *et al*., 2009; Zhang *et al*., 2012) (Supplementary Table S2). Thus, s-type genes existed as haplotypes in common wheat. However, both subgroups of s-type genes displayed significant differences in terms of gene composition and sequences, and thus might make different contributions to dough quality. Overall, evaluation of *Glu-B3* genes/haplotypes will enable development of haplotype-specific primers for marker-assisted selection.

Glu-D3 locus One s-type and seven m-type genes were identified at the *Glu-D3* locus from a single wheat genotype, which was by far the highest number of LMW-GS genes reported for this locus (Fig. 2c). Pseudogene *D3-586* was newly detected in common wheat, whereas the other seven genes have been investigated extensively (Supplementary Table S2 at *JXB* online), and covered all *Glu-D3* genes identified in wheat varieties, including Norin 61, Glenlea, Xiaoyan 54, Jing 411, and Aroona NILs (Supplementary Table S2) (Ikeda *et al*., 2002; Zhao *et al*., 2006, 2007; Huang and Cloutier, 2008; Dong *et al*., 2010; Zhang *et al*., 2011*a*, *b*). These *Glu-D3* genes were highly conserved, with only a few allelic variants (>99% identities), of which three novel active allelic variants (*D3-397*, *D3-444*, and *D3-522*) were detected with unique SNPs or indels in MCC, but functional analysis of these variants has been limited. Moreover, these *Glu-D3* genes shared >97% identities with LMW-GS genes isolated from *Aegilops tauschii* (Johal *et al*., 2004; Dong *et al*., 2010), which further confirmed the conservation of *Glu-D3* genes.

Relative genetic locations of LMW-GS genes at the Glu-3 loci

Typical LMW-GS genes were located at the *Glu-A3*, *Glu-B3*, and *Glu-D3* loci on the homoeologous group 1 chromosomes. However, little is known about the relative location of LMW-GS genes at individual loci due to the complexity of gene composition and the lack of appropriate methods of investigating this gene family. Recently, the recombination of 14 LMW-GS genes at *Glu-3* loci was analysed and the relative genetic position of these genes was determined (Dong *et al*., 2010). Subsequently, four more genes were detected and sequenced in Xiaoyan 54 (Zhang *et al*., 2011*a*, *b*). In the present study, based on the allelic relationship with the genes in Xiaoyan 54, all the LMW-GS genes in the MCC were located at a specific position in homoeologous group 1 chromosomes (Fig. 5). At the *Glu-A3* locus, two groups of LMW-GS genes, m_{AD} and i_A gene clusters, were found and little recombination was detected within groups, of which the i_A group was distal and the m_{AD} group was proximal to the centromere (Figs 2, 5). At the *Glu-B3* locus, although the relative position of *A3-548* was not determined, the m- and s-type genes might exist as two gene clusters (m_{BD} and s_{BD}). Also, m-type genes were more proximal to the centromere than s-type genes (Figs 2, 5). At the *Glu-D3* locus, tight linkage was identified only between *D3-441* and *D3-578a* or *D3-432* and *D3-578b* (Fig. 5), which could be explained by the close physical proximity (15.9 kb) of these genes (Dong *et al*., 2010). Additionally, *D3-385*, *D3-393*,

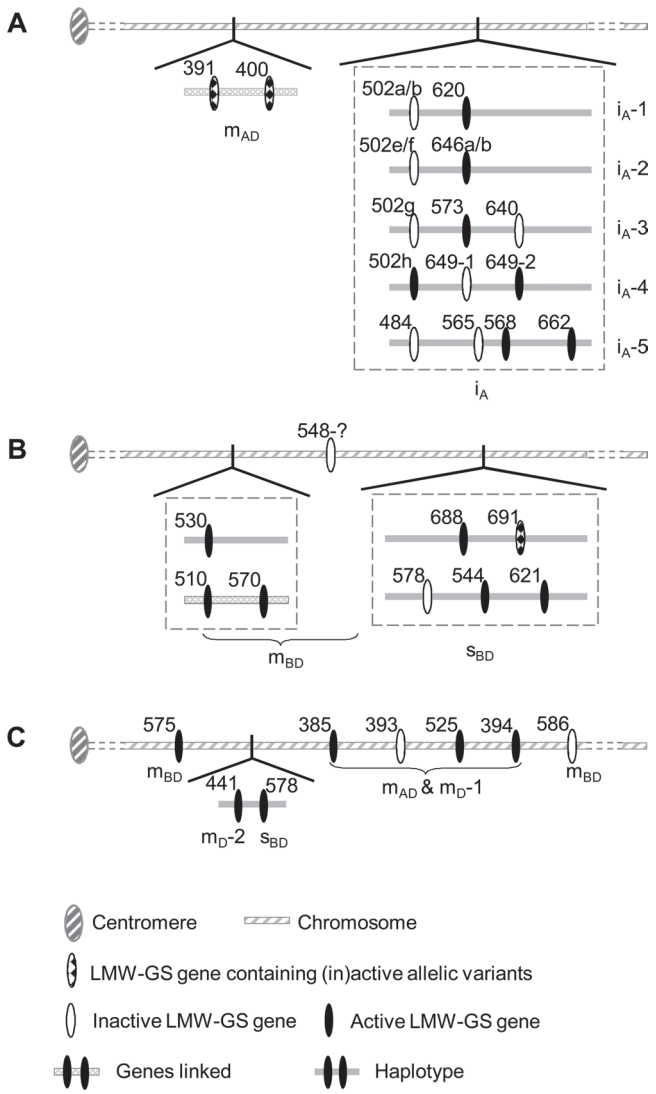

Fig. 5. Organization of the LMW-GS genes in homoeologous group 1 chromosomes. The relative locations of genes or haplotypes were determined based on Dong *et al*. (2010). The main allelic variants are displayed as representatives. The distances among genes do not represent the genetic or physical distances. (A) Relative genetic positions of LMW-GS genes at the *Glu-A3* locus. The i-type genes were tightly linked at the *Glu-A3* locus and formed five haplotype subgroups, and both m-type genes also generally co-segregated in the MCC. (B) Relative genetic positions of LMW-GS genes at the *Glu-B3* locus. The s-type genes were coupled at the *Glu-B3* locus and were of two principal haplotypes. (C) Relative genetic positions of the eight LMW-GS genes at the *Glu-D3* locus. Of these, only *D3-411* and *D3-578* were tightly linked.

D3-394, and *D3-525* genes were located in close proximity and had a high identity (Figs 4, 5). Based on the location and sequence analysis, LMW-GS genes that show high identity or belong to the same group may be tightly linked and located at the same position in the *Glu-3* loci.

In addition, twelve i-type haplotypes at the *Glu-A3* locus could be divided into five subgroups (i_A-1 to i_A-5; Fig. 5; Supplementary Fig. S12 at *JXB* online), and s-type haplotypes at the *Glu-B3* locus were divided into two subgroups, revealing the high diversity of the i- and s-type genes/haplotypes among wheat varieties. These subgroups were significantly different in terms of gene numbers and sequences, which suggests that the *Glu-A3* and *Glu-B3* loci in common wheat might be derived from several unique ancestors or have been involved in significant mutational or recombination events during the course of their evolution (Figs 1b, 5).

The complex LMW-GS gene family in Chinese wheat germplasm

The LMW-GS gene family was investigated using the MCC of Chinese wheat germplasm, which consists of 262 accessions with an estimated 70% genetic diversity compared with the full collection (Hao *et al*., 2011). Using the MCC, most (>15) LMW-GS genes were identified in individual wheat varieties. This allowed investigation of the classification and relationship of these genes in Chinese wheat germplasm.

The i-type genes were reported only at the *Glu-A3* locus (Zhang *et al*., 2004; Ikeda *et al*., 2006; Huang and Cloutier, 2008; Dong *et al*., 2010; Wang *et al*., 2010; Zhang *et al*., 2011*a*, *b*) or the A genome, for example *T. urartu* and *T. monococcum* (An *et al*., 2006; Ma *et al*., 2006; Caballero *et al*., 2008; Long *et al*., 2008). The findings also indicate that all i-type genes detected in MCC accessions were located at the *Glu-A3* locus (Fig. 4). Since i-type genes lacked the sequences encoding the N-terminal domain, m- and s-type genes without the N-terminal domain-coding sequences were used for the phylogenetic analysis. It was found that i-type genes had a closer relationship with s-type genes (s_{BD}) than m-type genes, excluding the m_D-2 group (Supplementary Fig. S13 at *JXB* online). This results confirmed that the i-type genes may be the result of a deletion event of s-type genes (Gao *et al*., 2007), and the i-type genes comprised a relatively young group of LMW-GS genes (Juhász and Gianibelli, 2006). The s-type genes were distributed at the *Glu-B3* and *Glu-D3* loci in common wheat and the progenitor of the wheat A genome, *T. urartu* (data not shown), but not at the *Glu-A3* locus in common wheat (Fig. 4). This suggests that their disappearance from the *Glu-A3* locus might be the result of elimination at the polyploid level. The m-type genes were common at the *Glu-3* loci, and the difference between the s- and m-types was not significant (D'Ovidio and Masci, 2004). The *D3-441* gene (m_D-2) and s-type genes were located at the same main branch, although they belonged to different groups (Fig. 4). These data confirm that the s-type genes probably originated from m-type genes due to mutation of MET to MEN in the N-terminal region (Masci *et al*., 1998; D'Ovidio and Masci, 2004). This also suggests that the m-type genes might be the oldest type of LMW-GS gene.

Genome sequence analysis of *Glu-3* loci revealed that both i-type and s-type genes/haplotypes existed together with the *Pm3* analogue and genetic marker *SFR159*, while most m-type genes, m_{AD}, m_D-1, and m_{BD}, were tightly linked with another genetic marker, *WHS179* (Wicker *et al*., 2003; Gao *et al*., 2007; Dong *et al*., 2010). Moreover, at the *Glu-A3* and *B3* loci, i-type and s-type genes were distal, while the m_{AD} and m_{BD} groups of genes were proximal to the centromere (Fig. 5) (Dong *et al*., 2010). At the *Glu-D3* locus, the s-type gene *D3-578* was also more distal to the centromere than the m_{BD} gene, *D3-575* (Fig. 5) (Dong *et al*., 2010). Thus, the phylogenetic analysis (Fig. 4; Supplementary Fig. S13 at *JXB* online), together with the linked genes/markers and their relative locations on chromosomes, suggest that i-type and s-type genes may have been derived from similar ancestral genes, and the m_{AD} (only *Glu-A3* genes) and m_{BD} groups of genes were orthologues among the A, B, and D subgenomes.

After sequence alignment and clustering analysis, the LMW-GS genes detected in MCC accessions were divided into six groups (Fig. 4). The genes at the *Glu-A3* locus were assigned to the i_A and m_{AD} groups, and those at the *Glu-B3* locus were divided into the s_{BD} and m_{BD} groups, whereas those at the *Glu-D3* locus were distributed widely among the $m_{D\text{-}1}$, $m_{D\text{-}2}$, s_{BD}, m_{BD}, and m_{AD} groups (Fig. 4). Thus, the *Glu-D3* genes showed higher diversity than those at the *Glu-A3* or *Glu-B3* loci in individual wheat varieties. The *Glu-A3* locus did not share the same group of LMW-GS genes as the *Glu-B3* locus (Figs 4, 5), which suggested that genes at the *Glu-A3* and *Glu-B3* loci evolved through different routes and had a distant evolutionary relationship. All genes at the *Glu-B3* locus and three *Glu-D3* genes comprised two groups (s_{BD} and m_{BD}); these homoeoalleles showed close relationships between the *Glu-B3* and *Glu-D3* loci, which were consistent with the derivation of BB and DD subgenomes from *Aegilops*.

Analysis of the LMW-GS gene family was performed using MCC accessions, consisting of foreign varieties, Chinese modern varieties, and landraces. The novel i-type haplotype, *A3-502h*/*A3-649-1*/*A3-649-2* was detected only in landraces and were absent from Chinese modern or foreign wheat varieties (Fig. 2a). This also occurred for genotypes containing *B3-624* and the genotype *B3-530*/*B3-548*/*B3-688*/*B3-691* (Fig. 2b). Although these haplotypes possessed an equal number of active genes to other haplotypes, they have not been selected for use in modern wheat breeding programmes. This may be because these genes or those linked to them have a detrimental effect on bread-making quality or yield potential. In contrast, *A3-394a*/*b*, *B3-601*/*604*, and the genotypes *D3-385*/*D3-376*/-/*D3-432*/-/*D3-575*/*D3-578a*/-, and 1BL/1RS lines were present only in modern or foreign varieties. The presence of these genes/genotypes in Chinese modern varieties might be the result of incorporation of foreign germplasm in breeding programmes in the past several decades. For example, the elite 1BL/1RS translocation lines were introduced into China in the 1970s and were exploited in modern wheat breeding since they contain several disease resistance and yield improvement genes. Also, the other genes/genotypes were probably introduced into Chinese wheat germplasm since they (or linked genes) increased the yield potential or bread-making quality.

Using the LMW-GS gene marker system and full-length gene cloning method, a representative population (MCC) was investigated, and the composition, organization, variation, and expression of LMW-GS genes were evaluated. Furthermore, the LMW-GS genes corresponding to all DNA fragments from the LMW-GS gene marker system were identified (Table 1; Supplementary S1 at *JXB* online). The expression profile of these genes was revealed by comparing the genomic DNA and cDNA data. These data will facilitate the update of the LMW-GS gene marker system which can be used to separate, identify, and characterize LMW-GS genes efficiently in common wheat.

Supplementary data

Supplementary data are available at *JXB* online.

Figure S1. Sequence alignments of the *A3-391* gene identified in the MCC.

Figure S2. Sequence alignments of the *A3-400* gene identified in the MCC.

Figure S3. Sequence alignments of the *A3-502* gene identified in the MCC.

Figure S4. Sequence alignments of the *B3-530* gene identified in the MCC.

Figure S5. Sequence alignments of the *B3-544* gene identified in the MCC.

Figure S6. Sequence alignments of the *B3-621* and *B3-688* genes identified in the MCC.

Figure S7. Sequence alignments of the *D3-441* gene identified in the MCC.

Figure S8. Sequence alignments of the *D3-525* gene identified in the MCC.

Figure S9. Sequence alignments of the *D3-578* gene identified in the MCC.

Figure S10. Sequence alignments of the *D3-586* gene identified in the MCC.

Figure S11. Sequence alignments of the deduced proteins of 15 representative LMW-GS genes from MCC accessions.

Figure S12. Phylogenetic reconstruction of all i-type LMW-GS genes and their allelic variants identified from MCC accessions.

Figure S13. Phylogenetic reconstruction of all LMW-GS genes and their allelic variants with removed sequences coding for N-terminal domains.

Table S1. Gene-specific primers used for cloning rare allelic variants.

Table S2. Nucleotide sequence identities of LMW-GS genes from MCC to the previously reported *Glu-A3*, *B3*, and *D3* alleles/genes.

Acknowledgements

We thank Professor Moshe Feldman (The Weizmann Institute of Science, Rehovot, Israel) for constructive suggestions on this work. This work was supported by the Ministry of Science and Technology of China (2009CB118300) and the Ministry of Agriculture of China for transgenic research (2013ZX08009-003-004 and 2013ZX08002-004).

References

An X, Zhang Q, Yan Y, *et al.*. 2006. Cloning and molecular characterization of three novel LMW-i glutenin subunit genes from cultivated einkorn (*Triticum monococcum* L.). *Theoretical and Applied Genetics* **113,** 383–395.

Appelbee MJ, Mekuria GT, Nagasandra V, Bonneau JP, Eagles HA, Eastwood RF, Mather DE. 2009. Novel allelic variants encoded at the *Glu-D3* locus in bread wheat. *Journal of Cereal Science* **49,** 254–261.

Bekes F, Kemeny S, Morell M. 2006. An integrated approach to predicting end-product quality of wheat. *European Journal of Agronomy* **25,** 155–162.

Branlard G, Dardevet M, Saccomano R, Lagoutte F, Gourdon J. 2001. Genetic diversity of wheat storage proteins and bread wheat quality. *Euphytica* **119,** 59–67.

Caballero L, Martin MA, Alvarez JB. 2008. Allelic variation for the high- and low-molecular-weight glutenin subunits in wild diploid wheat (*Triticum urartu*) and its comparison with durum wheats. *Australian Journal of Agricultural Research* **59,** 906–910.

Cassidy BG, Dvorak J, Anderson OD. 1998. The wheat low-molecular-weight glutenin genes: characterization of six new genes and progress in understanding gene family structure. *Theoretical and Applied Genetics* **96,** 743–750.

D'Ovidio R, Masci S. 2004. The low-molecular-weight glutenin subunits of wheat gluten. *Journal of Cereal Science* **39,** 321–339.

Dong LL, Zhang XF, Liu DC, *et al.*. 2010. New insights into the organization, recombination, expression and functional mechanism of low molecular weight glutenin subunit genes in bread wheat. *PLoS One* **5**: e13548.

Eagles HA, Hollamby GJ, Gororo NN, Eastwood RF. 2002. Estimation and utilisation of glutenin gene effects from the analysis of unbalanced data from wheat breeding programs. *Australian Journal of Agricultural Research* **53,** 367–377.

Gao S, Gu YQ, Wu J, *et al.*. 2007. Rapid evolution and complex structural organization in genomic regions harboring multiple prolamin genes in the polyploid wheat genome. *Plant Molecular Biology* **65,** 189–203.

Gupta RB, Bekes F, Wrigley CW. 1991. Prediction of physical dough properties from glutenin subunit composition in bread wheats—correlation studies. *Cereal Chemistry* **68,** 328–333.

Gupta RB, Paul JG, Cornish GB, Palmer GA, Bekes F, Rathjen AJ. 1994. Allelic variation at glutenin subunit and gliadin loci, *Glu-1, Glu-3* and *Gli-1*, of common wheat. 1. Its additive and interaction effects on dough properties. *Journal of Cereal Science* **19,** 9–17.

Hao C, Wang L, Ge H, Dong Y, Zhang X. 2011. Genetic diversity and linkage disequilibrium in Chinese bread wheat (*Triticum aestivum* L.) revealed by SSR markers. *PLoS One* **6,** e17279.

He ZH, Liu L, Xia XC, Liu JJ, Pena RJ. 2005. Composition of HMW and LMW glutenin subunits and their effects on dough properties, pan bread, and noodle quality of Chinese bread wheats. *Cereal Chemistry* **82,** 345–350.

Howitt CA, Gale KR, Juhász A. 2006. Diagnostic markers for quality. In: Wrigley CW, Békés F, Bushuk W, eds. *Gliadin and glutenin.*

The unique balance of wheat quality . St Paul, MN: AACCI Press, 333–361.

Huang XQ, Cloutier S. 2008. Molecular characterization and genomic organization of low molecular weight glutenin subunit genes at the *Glu-3* loci in hexaploid wheat (*Triticum aestivum* L.). *Theoretical and Applied Genetics* **116,** 953–966.

Ikeda TM, Araki E, Fujita Y, Yano H. 2006. Characterization of low-molecular-weight glutenin subunit genes and their protein products in common wheats. *Theoretical and Applied Genetics* **112,** 327–334.

Ikeda TM, Nagamine T, Fukuoka H, Yano H. 2002. Identification of new low-molecular-weight glutenin subunit genes in wheat. *Theoretical and Applied Genetics* **104,** 680–687.

Jackson EA, Holt LM, Payne PI. 1983. Characterization of high molecular-weight gliadin and low-molecular-weight glutenin subunits of wheat endosperm by two-dimensional electrophoresis and the chromosomal localization of their controlling genes. *Theoretical and Applied Genetics* **66,** 29–37.

Johal J, Gianibelli MC, Rahman S, Morell MK, Gale KR. 2004. Characterization of low-molecular-weight glutenin genes in *Aegilops tauschii*. *Theoretical and Applied Genetics* **109,** 1028–1040.

Juhász A, Gianibelli MC. 2006. Low-molecular-weight glutenin subunits: insights into this abundant subunit group present in glutenin polymers. In: Wrigley CW, Békés F, Bushuk W, eds. *Gliadin and glutenin. The unique balance of wheat quality* . St Paul, MN: AACCI Press, 171–212.

Kumar S, Nei M, Dudley J, Tamura K. 2008. MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. *Briefings in Bioinformatics* **9,** 299–306.

Liu L, Ikeda TM, Branlard G, *et al.*. 2010. Comparison of low molecular weight glutenin subunits identified by SDS–PAGE, 2-DE, MALDI-TOF-MS and PCR in common wheat. *BMC Plant Biology* **10,** 124.

Long H, Huang Z, Wei YM, Yan ZH, Ma ZC, Zheng YL. 2008. Length variation of i-type low-molecular-weight glutenin subunit genes in diploid wheats. *Russian Journal of Genetics* **44,** 429–435.

Ma ZC, Wei YM, Long H, Yan ZH, Baum B, Zheng YL. 2006. Characterization of low-molecular-weight i-type glutenin subunit genes from diploid wheat in relation to the gene family structure. *Molecular Biology* **40,** 897–906.

Masci S, D'Ovidio R, Lafiandra D, Kasarda DD. 1998. Characterization of a low-molecular-weight glutenin subunit gene from bread wheat and the corresponding protein that represents a major subunit of the glutenin polymer. *Plant Physiology* **118,** 1147–1158.

Payne PI. 1987. Genetics of wheat storage proteins and the effect of allelic variation on bread-making quality. *Annual Review of Plant Physiology and Plant Molecular Biology* **38,** 141–153.

Saghai-Maroof MA, Soliman KM, Jorgensen RA, Allard RW. 1984. Ribosomal DNA spacer-length polymorphisms in barley: mendelian inheritance, chromosomal location, and population dynamics. *Proceedings of the National Academy of Sciences, USA* **81,** 8014–8018.

Shewry PR. 2009. Wheat. *Journal of Experimental Botany* **60,** 1537–1553.

Shewry PR, Tatham AS, Barro F, Barcelo P, Lazzeri P, 1995. Biotechnology of breadmaking: unraveling and manipulating the multi-protein gluten complex. *Nature Biotechnology* **13,** 1185–1190.

Singh NK, Shepherd KW, Cornish GB. 1991. A simplified SDS–PAGE procedure for separating LMW subunits of glutenin. *Journal of Cereal Science* **14,** 203–208.

Wang LH, Li GY, Pena RJ, Xia XC, He ZH. 2010. Development of STS markers and establishment of multiplex PCR for *Glu-A3* alleles in common wheat (*Triticum aestivum* L.). *Journal of Cereal Science* **51,** 305–312.

Wang LH, Zhao XL, He ZH, Ma W, Appels R, Pena RJ, Xia XC. 2009. Characterization of low-molecular-weight glutenin subunit *Glu-B3* genes and development of STS markers in common wheat (*Triticum aestivum* L.). *Theoretical and Applied Genetics* **118,** 525–539.

Wicker T, Yahiaoui N, Guyot R, Schlagenhauf E, Liu ZD, Dubcovsky J, Keller B. 2003. Rapid genome divergence at orthologous low molecular weight glutenin loci of the A and A(m) genomes of wheat. *The Plant Cell* **15,** 1186–1197.

Zhang W, Gianibelli MC, Rampling LR, Gale KR. 2004. Characterisation and marker development for low molecular weight glutenin genes from *Glu-A3* alleles of bread wheat (*Triticum aestivum*. L). *Theoretical and Applied Genetics* **108,** 1409–1419.

Zhang XF, Liu DC, Jiang W, Guo XL, Yang WL, Sun JZ, Ling HQ, Zhang AM. 2011a. PCR-based isolation and identification of full-length low-molecular-weight glutenin subunit genes in bread wheat (*Triticum aestivum* L.). *Theoretical and Applied Genetics* **123,** 1293–1305.

Zhang XF, Jin H, Zhang Y, Liu DC, Li GY, Xia XC, He ZH, Zhang AM. 2012. Composition and functional analysis of low-molecular-weight glutenin alleles with Aroona near-isogenic lines of bread wheat. *BMC Plant Biology* **12,** 243.

Zhang XF, Liu DC, Yang WL, Liu KF, Sun JZ, Guo XL, Li YW, Wang DW, Ling HQ, Zhang AM. 2011b. Development of a new marker system for identifying the complex members of the low-molecular-weight glutenin subunit gene family in bread wheat (*Triticum aestivum* L.). *Theoretical and Applied Genetics* **122,** 1503–1516.

Zhao XL, Xia XC, He ZH, Gale KR, Lei ZS, Appels R, Ma W. 2006. Characterization of three low-molecular-weight *Glu-D3* subunit genes in common wheat. *Theoretical and Applied Genetics* **113,** 1247–1259.

Zhao XL, Xia XC, He ZH, Lei ZS, Appels R, Yang Y, Sun QX, Ma W. 2007. Novel DNA variations to characterize low molecular weight glutenin *Glu-D3* genes and develop STS markers in Common Wheat. *Theoretical and Applied Genetics* **114,** 451–460.